A nonparametric random-coefficients approach

Vermunt, J.K.; van Dijk, L.A.

Published in:
Multilevel Modelling Newsletter

Publication date:
2001

Paper published in Multilevel Modelling Newsletter, 2001, 13, 6-13

A non-parametric random-coefficient approach: the latent class regression model

Jeroen K. Vermunt and Liesbet A. van Dijk
Department of Methodology and Statistics, Tilburg University

1 Introduction

Most of the work in the area of random-coefficient modelling has focused on parametric methods in which the random coefficients are assumed to come from a known distribution, typically a multivariate normal distribution (see e.g. Bryk and Raudenbush, 1992; Goldstein, 1995; Hedeker and Gibbons, 1996; Agresti et al., 2000). This paper presents latent class (LC) analysis as a non-parametric random-coefficient model. Advantages of our proposed LC regression model are that less restrictive assumptions are made about the distribution of the random effects and that any model of the generalised linear modelling (GLM) family can be dealt with without increasing computation time. User friendly software with an SPSS-like interface is available to apply the proposed method (Vermunt and Magidson, 2000; www.LatentGold.com).

In the next section, we describe the LC regression model and compare it with the parametric random-coefficient model. Section three discusses parameter estimation by maximum likelihood (ML) and section four presents an application using an empirical data set. We end with some final remarks.

2 The latent class regression model

Let i denote a level-1 case within the level-2 case j. Let x denote a level-1 predictor and w a level-2 predictor. The general parametric level-2 model can be defined as follows:

$$\begin{aligned} \eta_{ij} &= \sum_{q=0}^{Q} \beta_{qj} x_{qij} + e_{ij}, \\ \beta_{qj} &= \sum_{s=0}^{S} \gamma_{qs} w_{sj} + u_{qj}, \end{aligned} \tag{1}$$

where $\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T})$. The distribution of e_{ij} can be any function belonging to the exponential family. Note that x_{0ij} and w_{0j} equal 1, which makes β_{0j} and γ_{q0} intercepts.

Using the same notation as in Equation (1) and indexing the latent classes by k, the LC regression model can be defined as follows:

$$\begin{aligned}\eta_{ij} &= \sum_{q=0}^{Q} \beta_{qk} x_{qij} + e_{ij}, \\ \beta_{qk} &= \sum_{s=0}^{S} \gamma_{qs} w_{sj} + u_{qk},\end{aligned} \tag{2}$$

where the distribution of $\mathbf{u}_k$ is unspecified, that is, $p(\mathbf{u}_k) = \pi_k$. For identification and comparability with the parametric two-level model, we set $\sum_{k=1}^{K} u_{qk}\pi_k = 0$. Note that in the standard formulation of the LC regression model, the first equation suffices.

Comparison of the LC model described in Equation (2) with the parametric two-level model of Equation (1) shows that rather than having a separate set of regression coefficients for each individual coming from a multivariate normal distribution, we assume that there exists a finite number of subgroups with different regression coefficients (Wedel and DeSarbo, 1994). This can be seen as a fundamental difference between the two models, especially if one is interested in identifying latent classes. However, the LC regression model can also be seen as a nonparametric two-level model; that is, as a two-level model in which no assumptions are made about the distributional form of the random effects. With the maximum number of identifiable latent classes, the distribution may be interpreted as a non-parametric distribution (Laird, 1978; Rabe-Hesketh et al., 2001). In practice, however, we will stop increasing the number of latent classes when the model fit does no longer improve. It should be noted that the current LC regression model cannot deal with more than two levels.

The conceptual equivalence between the LC regression and the two-level model becomes even clearer if we compute the second-order moments of the random coefficients from the standard latent class parameters. In a model without cross-level interactions, these are obtained by

$$\tau_{qq'} = \sum_{k=1}^{K} u_{qk} u_{q'k} \pi_k = \sum_{k=1}^{K} \left(\beta_{qk} - \gamma_{q0}\right) \left(\beta_{q'k} - \gamma_{q'0}\right) \pi_k, \tag{3}$$

where $\gamma_{q0} = \sum_{k=1}^{K} \beta_{qk}\pi_k$. Equation (3) shows that the results of a LC regression analysis can be summarised in the same way as of a two-level model; that is, in terms of a fixed and random part.

3 Parameter estimation

LC regression models are usually estimated by maximum likelihood (ML). The likelihood contribution of level-two unit j, equals

$$f(\mathbf{y}_j|\mathbf{x}_j, \mathbf{w}_j) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_j|\mathbf{x}_j, \mathbf{w}_j) = \sum_{k=1}^{K} \pi_k \prod_i f_k(y_{ij}|\mathbf{x}_j, \mathbf{w}_j),$$

where K is the number of latent classes and $f_k(y_{ij}|\mathbf{x}_j, \mathbf{w}_j)$ is a class-specific density. This density can be any function belonging to the exponential family.

The most popular algorithm to solve the ML estimation problem is the EM algorithm. The Latent GOLD software (Vermunt and Magidson, 2000) that was used for the example reported in the next section combines EM with Newton-Raphson. More precisely, the estimation process starts with a number of EM iterations and switches to Newton-Raphson when the relative change in the parameters is small. Local optima are avoided by using multiple sets of random starting values. Other software packages that can be used to estimate LC regression models are LEM (Vermunt, 1997), GLIMMIX (Wedel and DeSarbo, 1994), and GLLAMM (Rabe-Hesketh et al., 2001).

Contrary to the non-parametric method, parameter estimation in parametric random-coefficient models can become quite complex and time consuming when the distribution of the dependent variable is non-normal, such as with discrete response variables. Approximation methods to deal with the complicated integrals in the likelihood equations are numerical integration, Monte Carlo integration, and first- or second-order Taylor expansion of the link function (Agresti et al., 2000). It should be noted that the quite popular quadrature approximation of the likelihood that is used in the MIXOR (Hedeker and Gibbons, 1996) and GLLAMM (Rabe-Hesketh et al, 2001) packages is equivalent to using a LC model with many latent classes, where the location and weights of the classes are fixed rather than estimated from the data.

4 Application to attitudes towards abortion data

In order to compare the results of parametric and non-parametric random-coefficient models, we used a data set obtained from the data library of the Multilevel Models Project, at the Institute of Education, University of London (multilevel.ioe.ac.uk/intro/datasets.html). The data consist of 264 participants in 1983 to 1986 yearly waves from the British Social Attitudes Survey (McGrath and Waterton, 1986). It is a three-level data set: individuals are nested within constituencies and and time-points are nested within individuals. We will only make use of the latter nesting, which means that we are dealing with a standard repeated measures model. As was shown by Goldstein (1995), the highest level variance – between constituencies – is so small that it can reasonably be ignored.

The dependent variable is the number of yes responses on seven yes/no questions as to whether it is woman's right to have an abortion under a specific circumstance. Because this variable is a count with a fixed total, it most natural to work with a logit link and binomial error function. Individual level predictors in the data set are religion, political preference, gender, age, and self-assessed social class. In accordance with the results of Goldstein (1995), we found no significant effects of gender, age, self-assessed social class, and political preference. Therefore, we did not used these predictors in the further analysis. The predictors that were used are the level-1 predictor year of measurement (1=1983; 2=1984; 3=1985; 4=1986) and the level-2 predictor religion (1=Roman Catholic, 2=Protestant; 3=Other; 4=No religion).

The non-parametric models were estimated by means of version 2.0 of the Latent GOLD program (Vermunt and Magidson, 2000). Using the elementary statistics computations described in Equation (3), we obtained the multilevel type γ and τ parameters from the standard LC regression output. The parametric models were estimated with quadrature approximation of the likelihood. We used 10 nodes for the random intercept and 6 nodes for random slopes, which with 3 random slopes amounts to having a restricted "latent class" model with 2160 latent classes. The quadrature method was implemented in an experimental version of Latent GOLD. It is, however, not available in version 2.0 of the program.

First, three models without random effects were estimated: an intercept-only model (Ia), a model with a linear effect of year (Ib), and a model with year dummies (Ic). Models Ib and Ic also contained the nominal level-2 predictor religion. The test results reported in the first part of Table 1 show that year and religion have significant effects on the dependent variable and that it is better to treat year as non-linear.

[Insert Table 1 about here]

We proceeded by adding a random intercept to Model Ic using the parametric and non-parametric approach described in this paper (Models IIa-IIe). The test results show that both the parametric and the non-parametric random-effects models fit better than Model Ic. When using a latent-class approach, the model with 4 classes is the best one in terms of BIC value. It can also be seen that the 4-class model fits much better than the parametric model.

Subsequently, we included random slopes (Models IIIa-IIIe). Within the parametric approach, random slopes did not improve the fit in terms of BIC. In contrast, the LC models with random slopes are better than the models without random slopes. Again the 4-class model is the best one in terms of BIC. It turns out that this data set, the more flexible non-parametric approach is better able to capture the individual variation in the slopes than the more restricted parametric method, even with the same number of parameters as in the case of the 3-class model.

[Insert Table 2 about here]

Table 2 reports the multilevel parameter estimates for Models Ic, IIa, IId, IIIa, and IIId. As far as the fixed part is concerned, the substantive conclusions would be similar in all five models. The attitudes are most positive at the last time point (reference category) and most negative at the second time point. Furthermore, the effects of religion show that people without religion (reference category) are most in favour and Roman Catholics and Others are most against abortion. Protestants have a position that is close to the no-religion group. A difference between the parametric and non-parametric models is that in the former, Others are as extreme as Roman Catholics, while in the latter it is clearly an intermediate group

Also the random parts of the parametric models are quite similar. Some differences are that in the variance of the intercept is higher in Model IId than in Model IIa. The intercept, time-point one and time-point two variances are somewhat higher in Model IIIa than in Model IIId, but the time-point three variance is much lower. Furthermore, the covariances are much higher in the parametric than in the nonparametric model.

[Insert Figure about here]

The Figure depicts the random part of the 4-class model with random slopes (Model IIId) using standard latent class parameters. As can be seen, the 4 latent classes show different time patterns. The largest class 1 is most against abortion and class 3 is most in favour of abortion. Both latent classes are very stable over time. The overall level of latent class 2 is somewhat higher than of class 1, and it shows somewhat more change of the attitude over time. People belonging to latent class 4 are very instable: at the first two time points they are similar to class 2, at the third time point to class 4, and at the last time point again to class 2. Class 4 could therefore be labelled as random responders. It is interesting to note that in a three-class solution the random-responder class and class two are combined. Thus, by going from a three- to a four-class solution one identifies the interesting group with less stable attitudes.

5 Conclusions

In this paper we proposed using the LC regression model as a tool for random-coefficient modelling. We showed how to transform the standard LC regression parameters into multilevel parameters, yielding the same type of insight into the random structure as with a parametric random-coefficients model. The empirical example showed that the assumption of multivariate normality of the random coefficients may sometimes be too restrictive: the LC models fitted much better and detected the random slopes.

An important advantage of the non-parametric approach that was not mentioned yet is the much shorter computation time. Actually, the abortion example is a small problem for the Latent GOLD program: estimation of the largest model (IIIe) took only 3 seconds. In contrast, the estimation of the parametric model with 4 random coefficients took 18 minutes.

6 References

Agresti A., Booth, J.G., Hobert, J.P., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology,* **30**, 27-80.

Bryk, A.S., and Raudenbusch, S.W. (1992). *Hierarchical linear models: application and data analysis.* Newbury Park, CA: Sage Publications.

Goldstein, H. (1995) *Multilevel statistical models.* New York: Halsted Press.

Hedeker, D. and Gibbons. R.D. (1996). MIXOR: A computer program for mixed effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine,* **49**, 157-176.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association, 73*, 805-811.

McGrath, K and Waterton, J. (1986). *British social attitudes, 1983-1986 panel survey.* London: Social and Community Planning Research, Technical Report.

Rabe-Hesketh, S., Pickles, A. and A. Skrondal (2001). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter,* **13**, 17-23.

Vermunt, J.K. and Magidson, J. (2000). *Latent GOLD User's guide.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K. (1997). *LEM: A general program for the analysis of categorical data.* Users' manual. Tilburg University

Wedel, M. and DeSarbo, W. (1994). A review of recent developments in latent class regression models. In: R.P. Bagozzi (editor), *Advanced methods of marketing research*, pp. 352-388. Cambridge, Massachusetts: Blackwell Publishers.

Table 1. Test results for the estimated models with the attitudes towards abortion data

Model	Log-Lik.	BIC	Npar
I. No random effects			
a. empty model	-2308.6	4622.8	1
b. time linear	-2215.2	4458.4	5
c. time dummies	-2188.4	4415.8	7
II. Random intercept (time dummies)			
a. parametric (10 nodes)	-1711.8	3468.1	8
b. 2-class	-1754.7	3559.5	9
c. 3-class	-1697.4	3456.2	11
d. 4-class	-1689.5	3451.4	13
e. 5-class	-1689.5	3462.6	15
III. Random intercept and slope (time dummies)			
a. parametric (10, 6, 6, and 6 nodes)	-1695.7	3486.1	17
b. 2-class	-1745.4	3557.8	12
c. 3-class	-1682.7	3460.2	17
d. 4-class	-1656.7	3436.1	22
e. 5-class	-1645.2	3441.0	27

Table 2. Estimates of multilevel parameters for Models Ic, IIa, IId, IIIa, and IIId of Table 1.

Effect	Model Ic	Model IIa[1]	Model IId[2]	Model IIIa[1]	Model IIId[2]
Fixed part					
γ_0	1.50 (0.07)	1.97 (0.13)	*1.89*	2.23 (0.16)	*1.83*
Time					
γ_1 (1983)	-0.13 (0.08)	-0.16 (0.08)	-0.16 (0.08)	-0.35 (0.12)	*-0.13*
γ_2 (1984)	-0.55 (0.07)	-0.68 (0.08)	-0.67 (0.08)	-0.91 (0.11)	*-0.70*
γ_3 (1985)	-0.22 (0.08)	-0.27 (0.08)	-0.26 (0.08)	-0.34 (0.12)	*-0.15*
Religion					
γ_4 (Catholic)	-1.08 (0.10)	-1.07 (0.21)	-1.64 (0.25)	-1.24 (0.31)	-0.95 (0.17)
γ_5 (Protestant)	-0.38 (0.06)	-0.49 (0.19)	-0.22 (0.14)	-0.57 (0.17)	-0.23 (0.11)
γ_6 (Other)	-0.82 (0.08)	-1.12 (0.17)	-0.66 (0.17)	-1.24 (0.20)	-0.52 (0.18)
Random part					
τ_{00}		1.45	*2.05*	2.36	*2.03*
τ_{11}				0.25	*0.19*
τ_{22}				0.42	*0.26*
τ_{33}				0.31	*0.60*
τ_{01}				-0.59	*0.00*
τ_{02}				-0.82	*-0.28*
τ_{03}				-0.29	*-0.10*
τ_{12}				0.30	*0.20*
τ_{13}				0.01	*-0.11*
τ_{23}				0.14	*-0.03*

1. In the quadrature procedure one estimates the Choleski decomposition of T rather than T itself. Our procedure does therefore not yield standard errors for the τ parameters. Standard errors could, however, be obtained by the delta method.

2. We do not report standard errors for the (italicised) parameters, which are derived from the Latent GOLD output using Equation (3). These standard errors could, however, be obtained by the delta method. It should be noted that Latent GOLD provides standard errors, as well as two types of Wald tests for the standard LC regression parameters.